# Hadoop平台的安装\_V1.0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时间 | 版本 | 公司 | 作者 | 校验者 |
| 2020/02/15 | V1.0 | BOE | Huan.Wang |  |
|  |  |  |  |  |

目录

[Hadoop平台的安装\_V1.0 1](#_Toc32770728)

[一、 服务器基础配置 2](#_Toc32770729)

[1、安装环境 2](#_Toc32770730)

[2、虚拟机硬件规划 2](#_Toc32770731)

[3、虚拟机安装软件版本 2](#_Toc32770732)

[4、配置机器的IP 2](#_Toc32770733)

[5、修改主机名hostname 4](#_Toc32770734)

[6、修改主机名和IP的映射关系 4](#_Toc32770735)

[7、关闭防火墙 4](#_Toc32770736)

[8、CentOS7中添加新用户并授权 6](#_Toc32770737)

[9、安装JDK 7](#_Toc32770738)

[10、安装和配置启动NTP服务（时间同步master节点） 8](#_Toc32770739)

[11、克隆虚拟机(node-2 和node-3) 10](#_Toc32770740)

[12、配置SSH免密登录 13](#_Toc32770741)

[13、安装mysql数据库 14](#_Toc32770742)

[14、设置slave到master时间同步（在slave节点操作）： 16](#_Toc32770743)

[二、Cloudera Manager安装 17](#_Toc32770744)

[1、主节点解压安装 17](#_Toc32770745)

[4、 相关启动脚本 18](#_Toc32770746)

[5、Cloudera Manager 安装 18](#_Toc32770747)

[三、HDFS服务安装 24](#_Toc32770748)

[1、HDFS安装 24](#_Toc32770749)

[2、HDFS安装测试 28](#_Toc32770750)

## 服务器基础配置

### 1、安装环境

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 操作系统 | 硬盘 | CPU | 内存 | 虚拟机 | Shell终端 |
| Win 10 | 256固态 | i7 8核 | 16G | Vmware 15.0 | SecureCRT7.0.4 |

### 2、虚拟机硬件规划

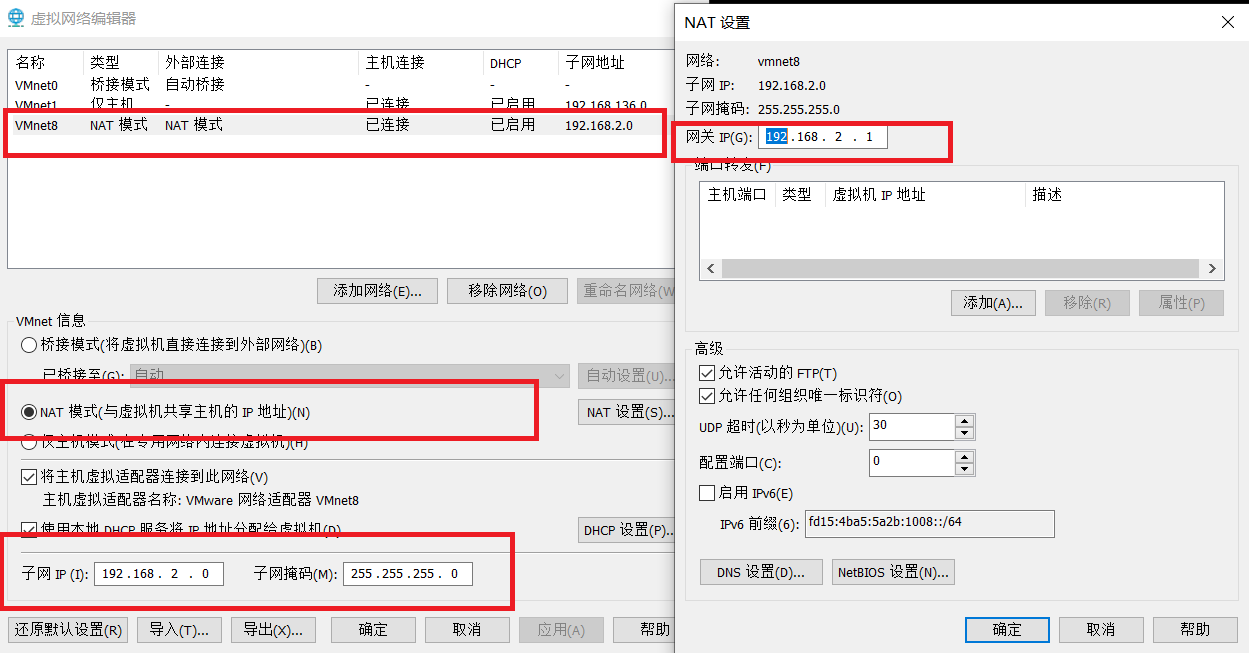
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 主机名 | IP | 内存 | 硬盘 | Root密码 | 普通用户 |
| node-1(master) | 192.168.2.100 | 4G | 30G | hadoop | **hadoop**:**wanghuan2054** |
| node-2(slave) | 192.168.2.101 | 4G | 30G | hadoop | **hadoop**:**wanghuan2054** |
| node-3(slave) | 192.168.2.102 | 4G | 30G | hadoop | **hadoop**:**wanghuan2054** |

### 3、虚拟机安装软件版本

|  |  |
| --- | --- |
| 软件名称 | 版本 |
| 操作系统 | CentOS-7.7-x86\_64-DVD-1503.iso |
| JDK | jdk-8u60-linux-x64.tar.gz |
| 数据库 | MySQL 5.7 |
| JDBC | mysql-connector-java-5.1.27-bin.jar |
| Cloudera Manager | 5.14.0 |
| CDH | 5.14.0 |

### 4、配置机器的IP

VMWARE网络设置 ， NAT方式，网关：192.168.2.0



centos7配置IP地址 ， centos7不再使用ifconfig命令，可通过命令 ip addr查看，修改IP两种方式

**第一种**：通过Linux图形界面进行修改（需要启动Centos图形化界面）

进入Linux图形界面 -> 右键点击右上方的两个小电脑 -> 点击Edit connections -> 选中当前网络System eth0 -> 点击edit按钮 -> 选择IPv4 -> method选择为manual -> 点击add按钮 -> 添加IP：192.168.2.100 子网掩码：255.255.255.0 网关：192.168.2.1 -> apply

**第二种**：修改配置文件方式（静态IP地址）

设置静态IP地址与动态IP差不多，也是要修改网卡配置文件 vi /etc/sysconfig/network-scripts/ifcfg-ens32    (最后一个为网卡名称)

（1）bootproto=static

（2）onboot=yes

（3）在最后加上几行，IP地址、子网掩码、网关、dns服务器

IPADDR=192.168.2.100

NETMASK=255.255.255.0

GATEWAY=192.168.2.1

DNS1=192.168.2.1 （默认DNS1与网关IP一致）

DNS2=8.8.8.8

（4）重启网络服务

systemctl restart network

ping www.baidu.com

### 5、修改主机名hostname

（1） vi /etc/sysconfig/network

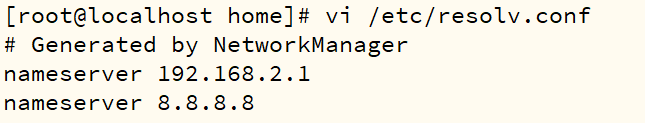
NETWORKING=yes

NETWORKING\_IPV6=no

HOSTNAME=node-1

GATEWAY=192.168.2.1

(2) cat /etc/resolv.conf (查看域名服务器是否变更)



同理， 其它机器全部修改，只需要把IP、和 hostname改一下。

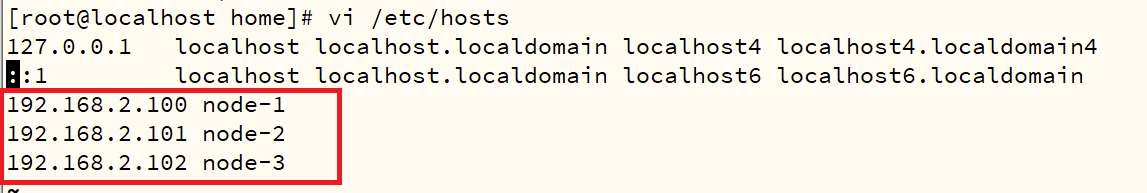
### 6、修改主机名和IP的映射关系

1) vi /etc/hosts

192.168.2.100 node-1

192.168.2.101 node-2

192.168.2.102 node-3

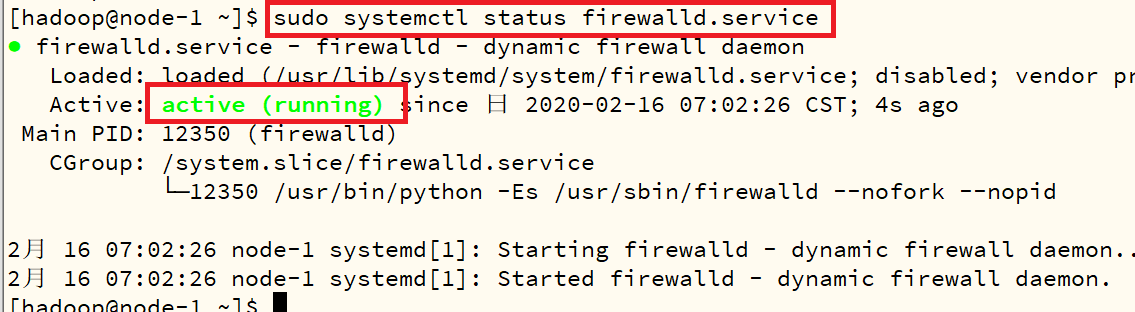


### 7、关闭防火墙

Centos7 默认使用防火墙firewalld，如果使用iptables需要重新配置一下。

1) 查看防火墙状态

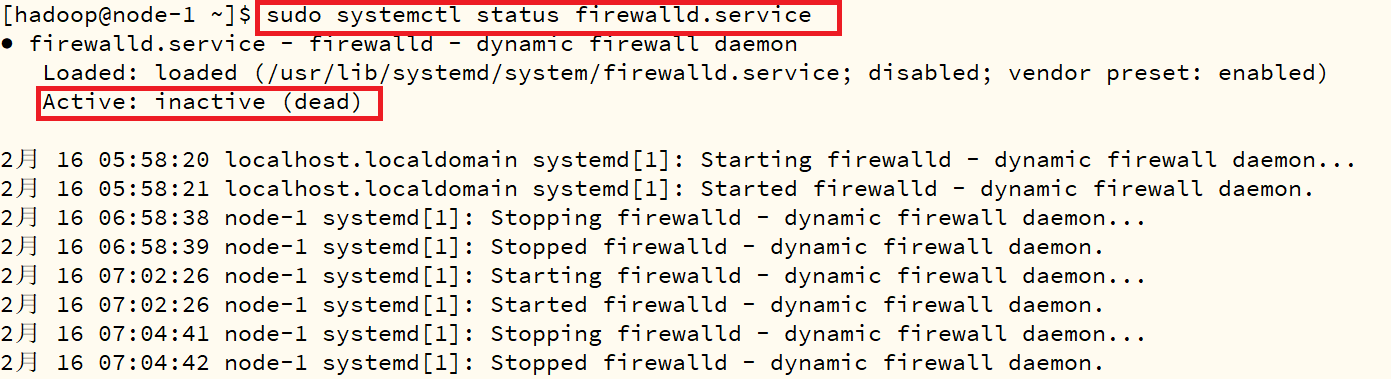
systemctl status firewalld.service



2) 关闭防火墙

systemctl stop firewalld.service  #停止firewall

systemctl disable firewalld.service #禁止firewall开机启动



3) 启动防火墙

systemctl start firewalld.service #启动firewall

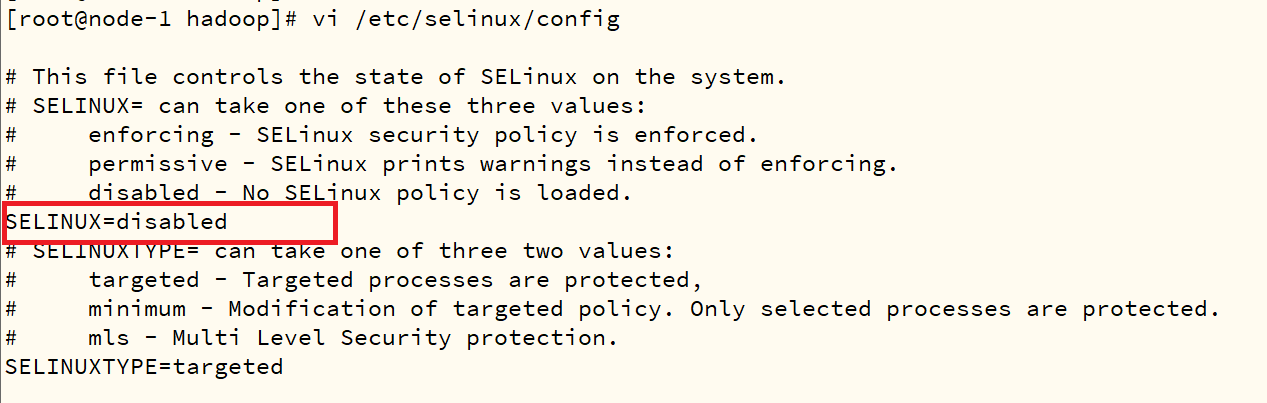
systemctl enable firewalld.service #允许firewall开机启动

4) 重启防火墙

systemctl restart firewalld.service

5) 关闭SELunx

vi /etc/selinux/config文件，将SELINUX的值改成disabled  #关闭SELinux开机自启



### 8、CentOS7中添加新用户并授权

1) hadoop用户

adduser hadoop

2) 设置密码

wanghuan2054

3) 新创建的用户并不能使用sudo命令，需要给他添加授权，添加hadoop用户到sudoers文件中

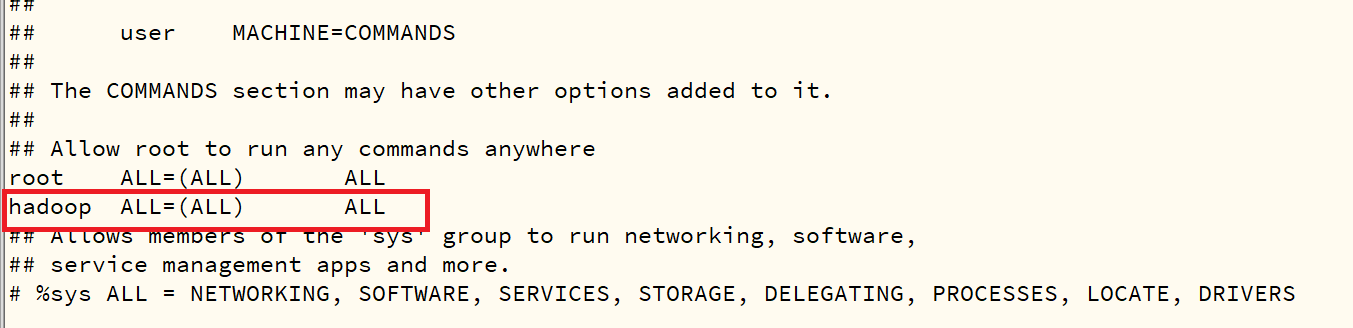
1. 添加sudoers文件可写权限，此文件默认没有w权限，需添加w权限后才能修改，改完后记得将w权限收回

**chmod -v u+w /etc/sudoers 赋写权限**

**chmod -v u-w /etc/sudoers 收回写权限**

2. 修改sudoers文件，**vi /etc/sudoers**

   在sudoers文件中找到如下位置并添加如下内容：**[用户名]    ALL=(ALL)    ALL**（如需新用户使用sudo时不用输密码，把最后一个**ALL**改为**NOPASSWD:ALL**即可）

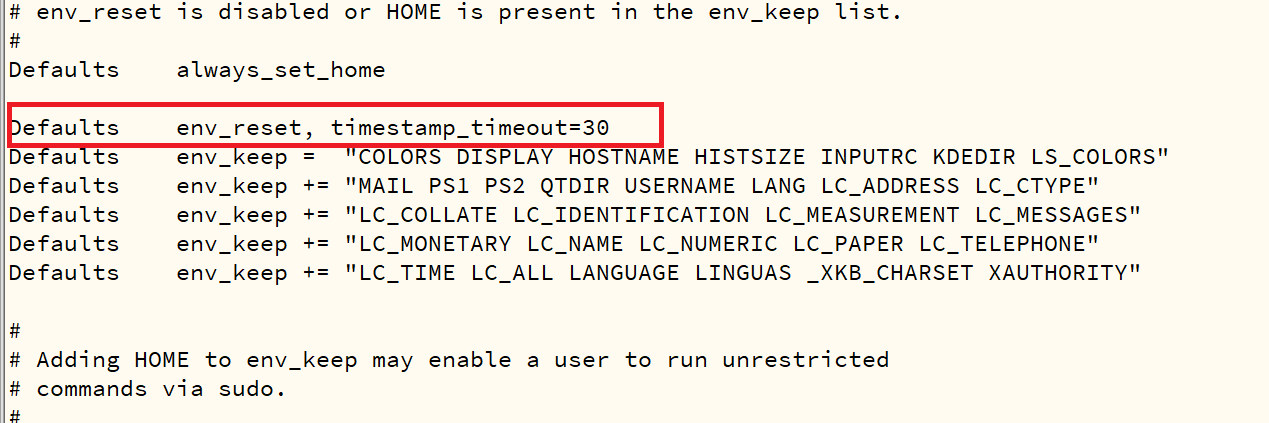


或者设置记住sudo密码时间，避免多次重复输入密码：

Defaults env\_reset改为：

Defaults env\_reset , timestamp\_timeout=30

#30就是代表时间，你可以设置为 10 或 30 ，表示10分钟或半小时。



### 9、安装JDK

yum -y install lrzsz , 用于上传下载文件的工具

rz 上传文件

sz 文件名 , 下载文件

1) rpm -qa | grep java

先查询是否存在open jdk环境

2) 若存在，卸载

rpm -e --nodeps 包名

3) 安装JDK

mkdir /usr/java

cd /usr/java/

rz // 上传jdk-8u60-linux-x64.tar.gz到 /usr/java目录下

tar -zxvf jdk-8u60-linux-x64.tar.gz -C .

编辑环境变量

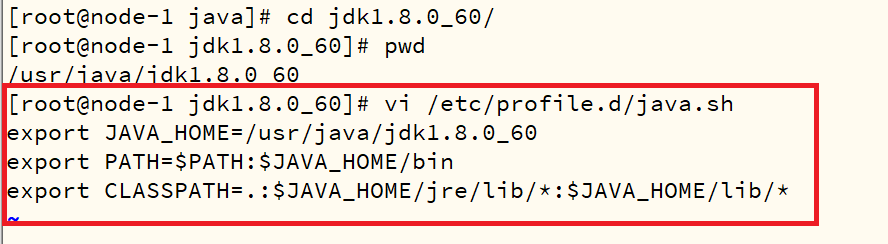
vi /etc/profile.d/java.sh

输入：

export JAVA\_HOME=/usr/java/jdk1.8.0\_60

export PATH=$PATH:$JAVA\_HOME/bin

export CLASSPATH=.:$JAVA\_HOME/jre/lib/\*:$JAVA\_HOME/lib/\*



使配置文件生效

source /etc/profile.d/java.sh

测试java安装是否成功

java -version

### 10、安装和配置启动NTP服务（时间同步master节点）

1) 安装 ntp服务

yum -y install ntp

2) 设置ntp开机启动 如下：

systemctl restart ntpd = systemctl restart ntpd.service

systemctl enable ntpd.service

systemctl start ntpd.service // 启动ntp

systemctl status ntpd.service //查看ntp服务

systemctl stop ntpd.service //关闭ntp服务

systemctl restart ntpd.service //重启ntp

3) 手动同步时间

ntpdate -u 1.cn.pool.ntp.org

4) 自动同步时间

在server节点上设置其ntp服务器为其自身，同时设置可以接受连接服务的客户端，是通过更改/etc/ntp.conf文件来实现，其中server为默认，新增加一个 restrict 段为可以接受服务的网段

restrict 127.0.0.1

restrict ::1

# Hosts on local network are less restricted.

# 允许内网其他机器同步时间

restrict 192.168.2.0 mask 255.255.255.0 nomodify notrap

# 中国这边最活跃的时间服务器 :

server 0.cn.pool.ntp.org

server 1.cn.pool.ntp.org

server 2.cn.pool.ntp.org

server 3.cn.pool.ntp.org

# allow update time by the upper server

# 允许上层时间服务器主动修改本机时间

restrict 0.cn.pool.ntp.org nomodify notrap noquery

restrict 1.cn.pool.ntp.org nomodify notrap noquery

restrict 2.cn.pool.ntp.org nomodify notrap noquery

restrict 3.cn.pool.ntp.org nomodify notrap noquery

# 外部时间服务器不可用时，以本地时间作为时间服务

server 127.0.0.1 # local clock

fudge 127.0.0.1 stratum 10

5）重启ntpd服务

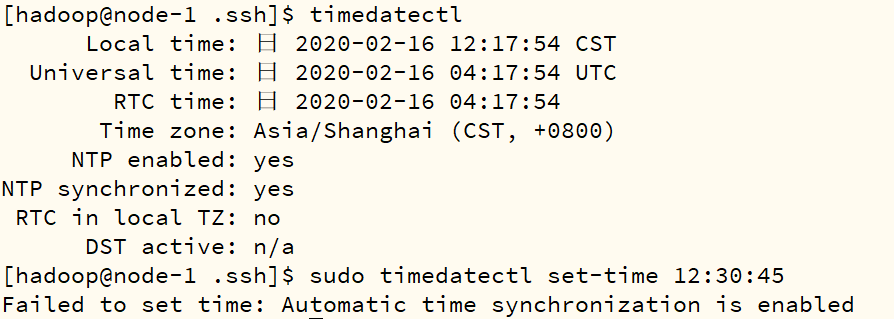
安装完成后设置ntp开机启动并启动ntp，如下：

systemctl enable ntpd

systemctl restart ntpd = systemctl restart ntpd.service

启用ntpd后，服务器就开启了ntpd自动同步，无法使用 timedatectl set-time HH:MM:SS重新设置时间。

如下：



参考文档：

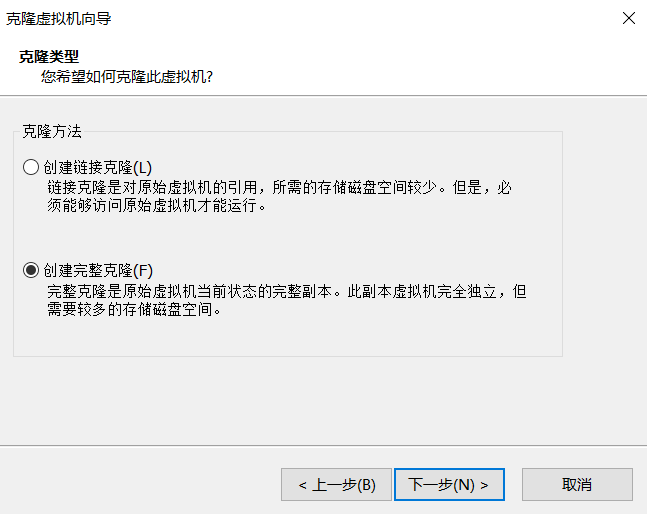
1. <https://www.jianshu.com/p/3701fb9f263b>

2. https://blog.csdn.net/qq\_38591756/article/details/85243965

### 11、克隆虚拟机(node-2 和node-3)

将master节点虚拟机关闭，进行可控，步骤如下（省略一些默认步骤）：

1. 选择创建完整克隆

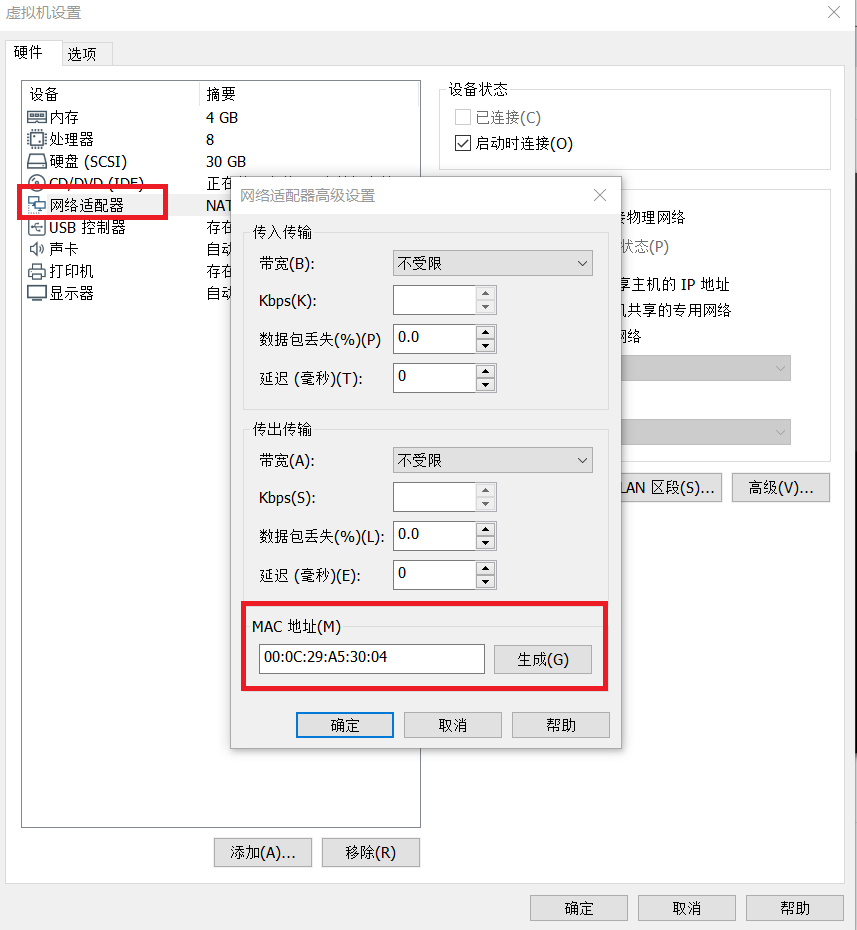


1. 填写虚拟机名称（和hostname保持一致）和保存位置



1. 重新生成mac地址

点击生成node-1虚拟机的mac地址



1. 修改hostname
2. vi /etc/sysconfig/network

NETWORKING=yes

NETWORKING\_IPV6=no

HOSTNAME=node-2

GATEWAY=192.168.2.1

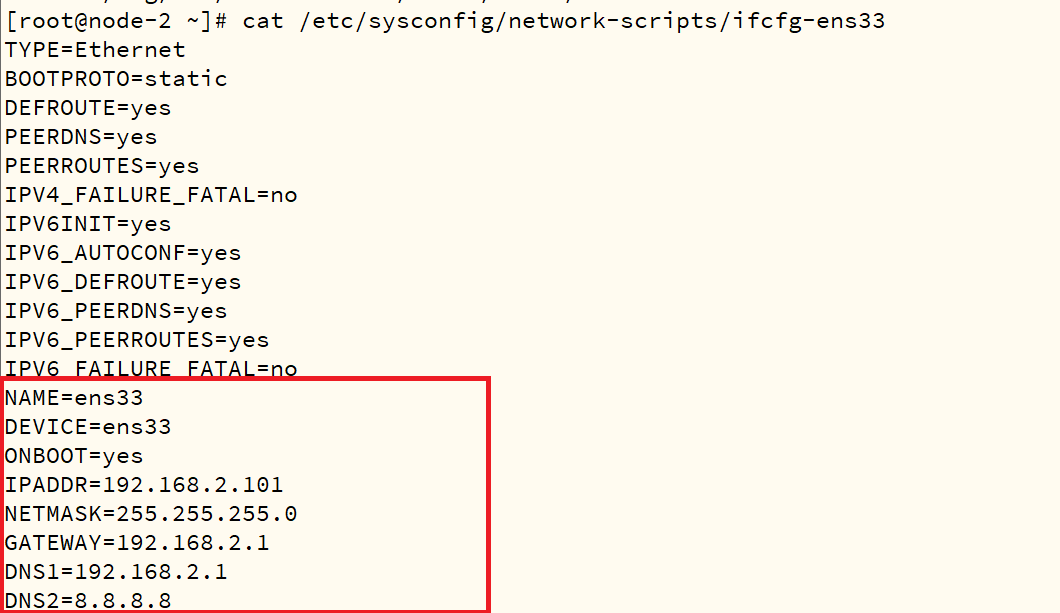
1. 修改IP，删除UUID

克隆完之后的网卡名称为ifcfg-eno16777736 ， ip addr查询当前网卡名为ifcfg-ens33

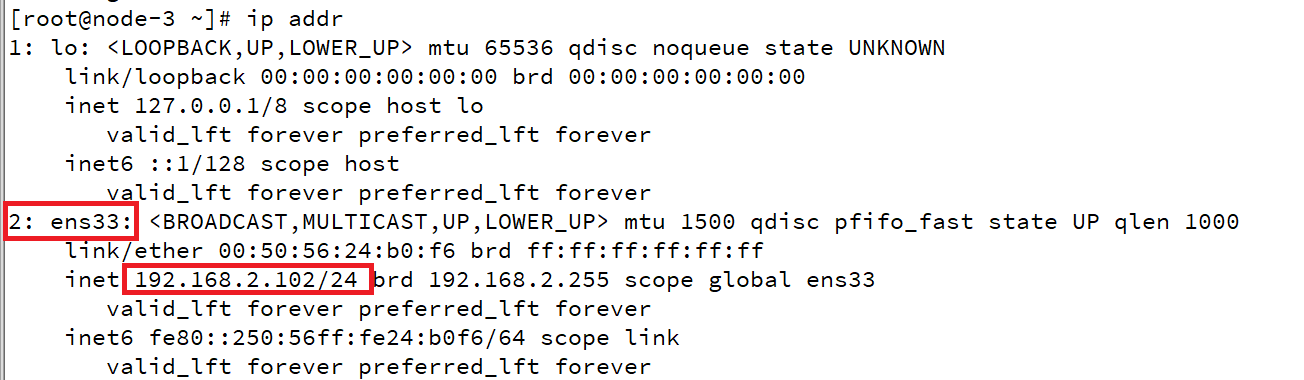
修改网卡名：

mv ifcfg-eno16777736 ifcfg-ens33

修改配置文件：vi /etc/sysconfig/network-scripts/ifcfg-ens33



修改成功之后：



1. 删除克隆机设备文件

rm -rf /etc/udev/rules.d/70-persistent-ipoib.rules

node-3执行同样的操作

### 12、配置SSH免密登录

1) 在node-1节点，进入~/.ssh目录，若存在该目录 ， 先删除

输入命令生成公钥密钥文件：ssh-keygen -t rsa，一路回车

node-2,node-3节点指定同样操作

2) 注意：确保.ssh文件夹的权限为700，id\_rsa.pub和authorized\_keys权限为644,其他权限都会导致SSH出问题。

3) 在node-1节点上

输入命令：ssh-copy-id node-1

再输入： ssh-copy-id node-2

输入命令：ssh-copy-id node-3

**在其它节点重复步骤3**

### 13、安装mysql数据库

Mysql账号密码：**root: 123456**

1）卸载CentOs自带的Maridb数据库

rpm -qa | grep mariadb

2）删除搜索出来的软件包名

rpm -e 搜索出来的软件包名

rpm -e --nodeps 强制删除

3）删除yum中有关maridb的相关软件包

yum -y remove mari\*

4）如果之前用的是yum方法安装的mysql但出了问题想重新安装，建议删干净，以免后续出现什么问题

rpm -qa | grep mysql

yum remove mysql-\*

find / -name mysql

rm -rf 搜索出来的文件名（例如 rm -rf /usr/local/mysql）

5） 安装MySQL

yum install mysql-community-server

systemctl start mysqld //启动MySQL服务

systemctl stop mysqld // 停止MySQL服务

systemctl status mysqld //查看MySQL服务

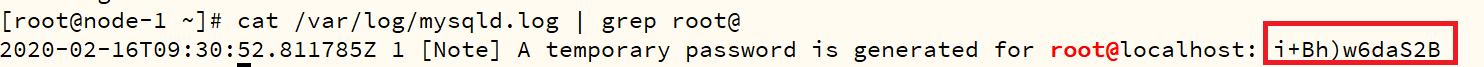
systemctl enable mysqld //设置开机启动

systemctl daemon-reload // 重新加载服务的配置文件

6） 修改root本地登录密码

mysql安装完成之后，在/var/log/mysqld.log文件中给root生成了一个默认密码。通过下面的方式找到root默认密码，然后登录mysql进行修改：

1. cat /var/log/mysqld.log | grep root@



1. mysql -uroot -pi+Bh)w6daS2B
2. set password for 'root'@'localhost'=password('password!');

mysql5.7默认的密码设置强度很高，降低密码长度和复杂度需要设置两个参数，具体参考文章https://www.cnblogs.com/kingsonfu/p/9803418.html

必须修改两个全局参数

首先，修改validate\_password\_policy参数的值

set global validate\_password\_policy=0;

这样就去除了规则 1，只有长度限制了

接下来修改参数 validate\_password\_length

mysql> select @@validate\_password\_length;#查看参数值

+----------------------------+

| @@validate\_password\_length |

+----------------------------+

| 8 |

+----------------------------+

set global validate\_password\_length=1;

mysql> select @@validate\_password\_length;#查看参数值

+----------------------------+

| @@validate\_password\_length |

+----------------------------+

| 4 |

+----------------------------+

只要设置的值小于4都为4。

这样就可以修改简单的密码了

update user set authentication\_string=password('123456') where user='root';

flush privileges;

退出：quit

重启服务：systemctl restart mysqld.service

注意点: 上述方法只在当前sql会话中有效， MySQL 5.7 关闭validate\_password密码策略永久有效：

如果不需要密码策略，在my.cnf文件中添加如下配置禁用即可：

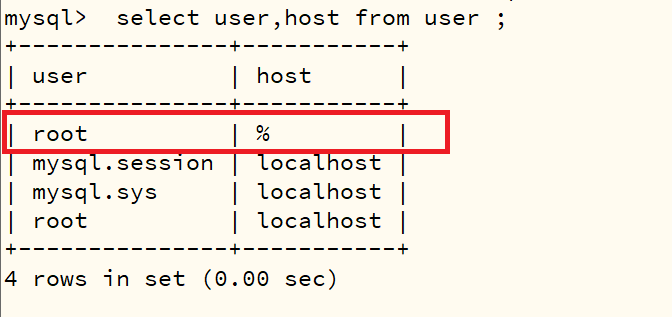
validate\_password = off （https://cloud.tencent.com/developer/article/1368306）

1. 允许远程连接

修改host为% ， 即在任何客户端地址下都能使用root用户登录

grant all on \*.\* to root@'%' identified by 'password' with grant option;

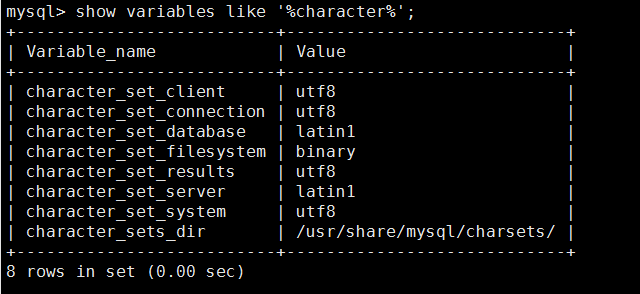
flush privileges;



7） 修改字符集为UTF-8

通过命令mysql -uroot -p 进入Mysql，注意是新密码，在mysql环境下通过如下命令查看字符集

show variables like '%character%';



我们看到上面的字符集不都是utf-8，现在我们在bash终端打开mysql配置文件

vi /etc/my.cnf

添加以下配置项

character-set-server=utf8

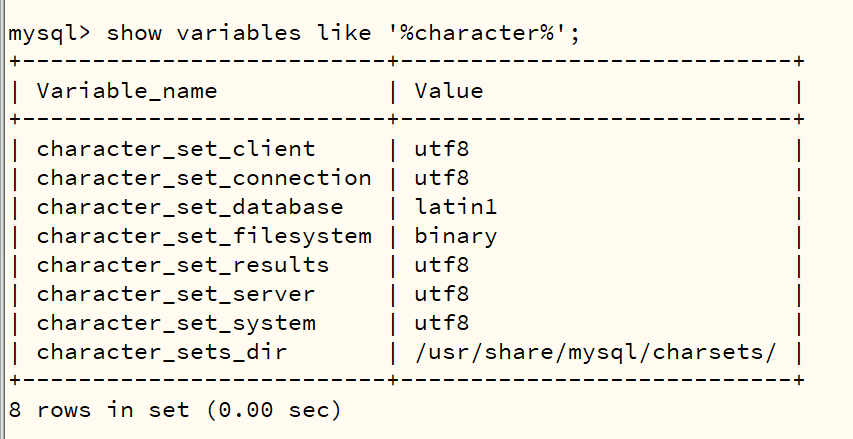
[client]

default-character-set=utf8

[mysql]

default-character-set=utf8

重启服务器，查看字符集有没有修改成功



### 14、设置slave到master时间同步（在slave节点操作）：

1. vi /etc/ntp.conf

内容如下：

# 让NTP Server master为内网的ntp服务器

server 192.168.2.100

fudge 192.168.2.100 nomodify notrap noquery

server 127.0.0.1 # local clock

fudge 127.0.0.1 stratum 10

# 不允许来自公网上ipv4和ipv6客户端的访问

restrict -4 default kod notrap nomodify nopeer noquery

restrict -6 default kod notrap nomodify nopeer noquery

# Local users may interrogate the ntp server more closely.

restrict 127.0.0.1

restrict ::12.

2. 重启ntpd服务：

systemctl ntpd restart

3. 手动同步：

ntpdate -u node-1 (首次时间同步需要执行)

4. 所有节点重启服务器：

reboot

重启后可以使用ntpq -p 查看网络中的NTP服务器

使用ntpstat 命令查看时间同步状态，但是需要等待几分钟才回看到同步信息 ， slave节点需要等待一段时间才开始同步

### 15、创建数据库

1. --hive

create database hive DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci ;

1. --activity monitor

create database amon DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci ;

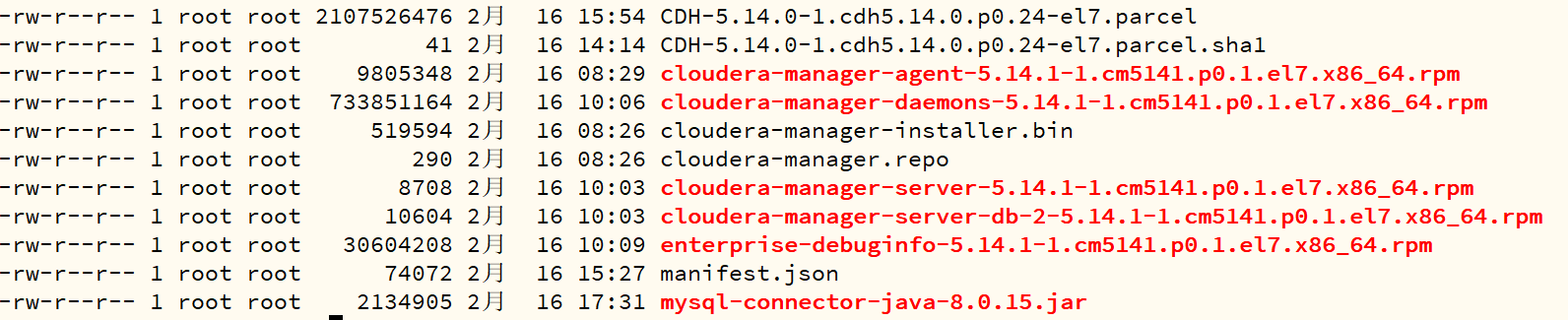
## Cloudera Manager安装

两种安装方式: cm-tar包，cm-rpm包，本安装采用rpm包

### 1、主节点安装

1. 上传CDH相关包到/opt/software下

mkdir /opt/software

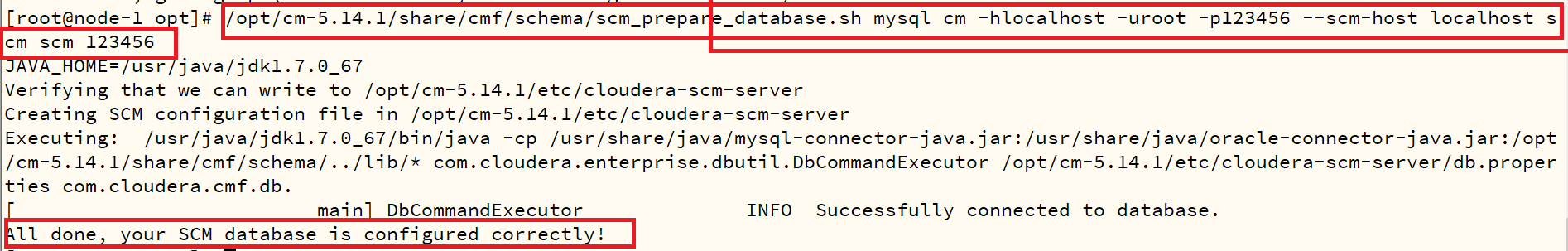


上传上述包，建议离线下载好安装

1. Copy mysql包到 /opt/cm-5.14.1/share/cmf/lib/下
2. 建立CM数据库

在主节点初始化CM5的数据库：账号密码：scm:123456

/opt/cm-5.14.1/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hlocalhost -uroot -p123456 --scm-host localhost scm scm 123456



create database hue DEFAULT CHARACTER SET utf8;

grant all on hue.\* TO 'hue'@'%' IDENTIFIED BY '123456';

create database oozie DEFAULT CHARACTER SET utf8;

grant all on oozie.\* TO 'oozie'@'%' IDENTIFIED BY '123456';

原文链接：https://blog.csdn.net/whq12789/article/details/88059990

1. 准备Parcels，用以安装CDH5

将CHD5相关的Parcel包放到主节点的/opt/cloudera/parcel-repo/目录中

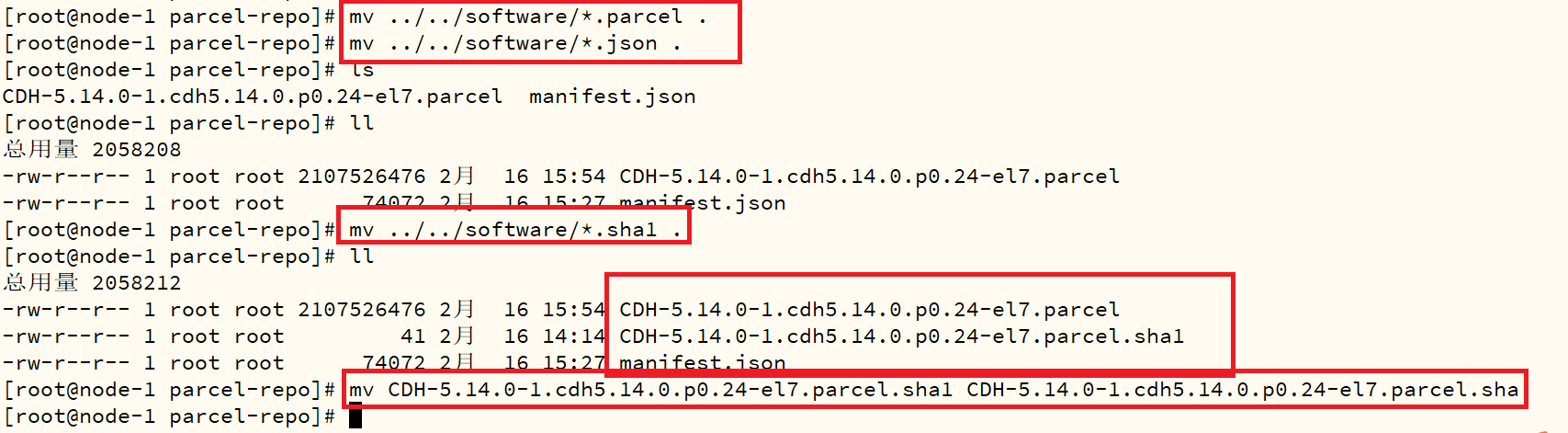
相关的文件如下：

CDH-5.14.1-1.cdh5.14.1.p0.2-el7.parcel

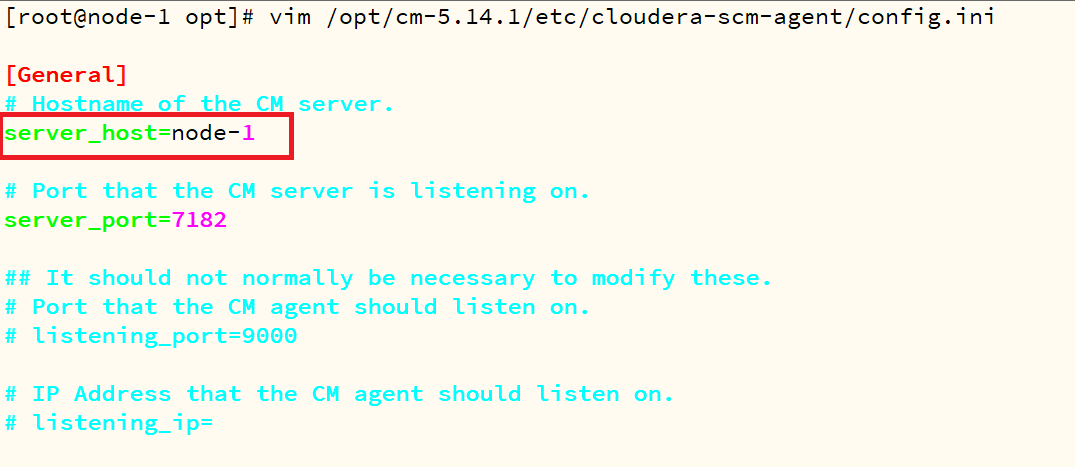
CDH-5.14.1-1.cdh5.14.1.p0.2-el7.parcel.sha1

manifest.json

最后将CDH-5.14.1-1.cdh5.14.1.p0.2-el7.parcel.sha1，重命名为CDH-5.14.1-1.cdh5.14.1.p0.2-el7.parcel.sha，这点必须注意，否则，系统会重新下载CDH-5.14.1-1.cdh5.14.1.p0.2-el7.parcel。



修改/opt/cm-5.14.1/etc/cloudera-scm-agent/config.ini中的server\_host为主节点的主机名。



1. 将/opt/cm-5.14.1目录复制到另外两个节点/opt目录

同步Agent到其他节点

scp -r /opt/cm-5.14.1 node-2:/opt

scp -r /opt/cm-5.14.1 node-3:/opt

在所有节点创建cloudera-scm用户

useradd --system --home=/opt/cm-5.14.1/run/cloudera-scm-server/ --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm

**所有节点创建parcels目录（所有节点）**

sudo mkdir -p /opt/cloudera/parcels

sudo chown cloudera-scm:cloudera-scm /opt/cloudera/parcels

解释：Clouder-Manager将CDHs从主节点的/opt/cloudera/parcel-repo目录中抽取出来，分发解压激活到各个节点的/opt/cloudera/parcels目录中

**原文链接：https://blog.csdn.net/silentwolfyh/article/details/54893826**

### 6、 启动脚本

6.1 node-1 启动cm server

启动服务端

/opt/cm-5.14.1/etc/init.d/cloudera-scm-server start

查看server启动的日志，是否有报错信息

tail -f /opt/cm-5.14.1/log/cloudera-scm-server/cloudera-scm-server.log

6.2 所有节点启动cm agent

启动Agent服务

/opt/cm-5.14.1/etc/init.d/cloudera-scm-agent start

查看agent启动的日志，是否有报错信息

tail -f /opt/cm-5.14.1/log/cloudera-scm-agent/cloudera-scm-agent.log

我们启动的其实是个service脚本，需要停止服务将以上的start参数改为stop就可以了，重启是restart。

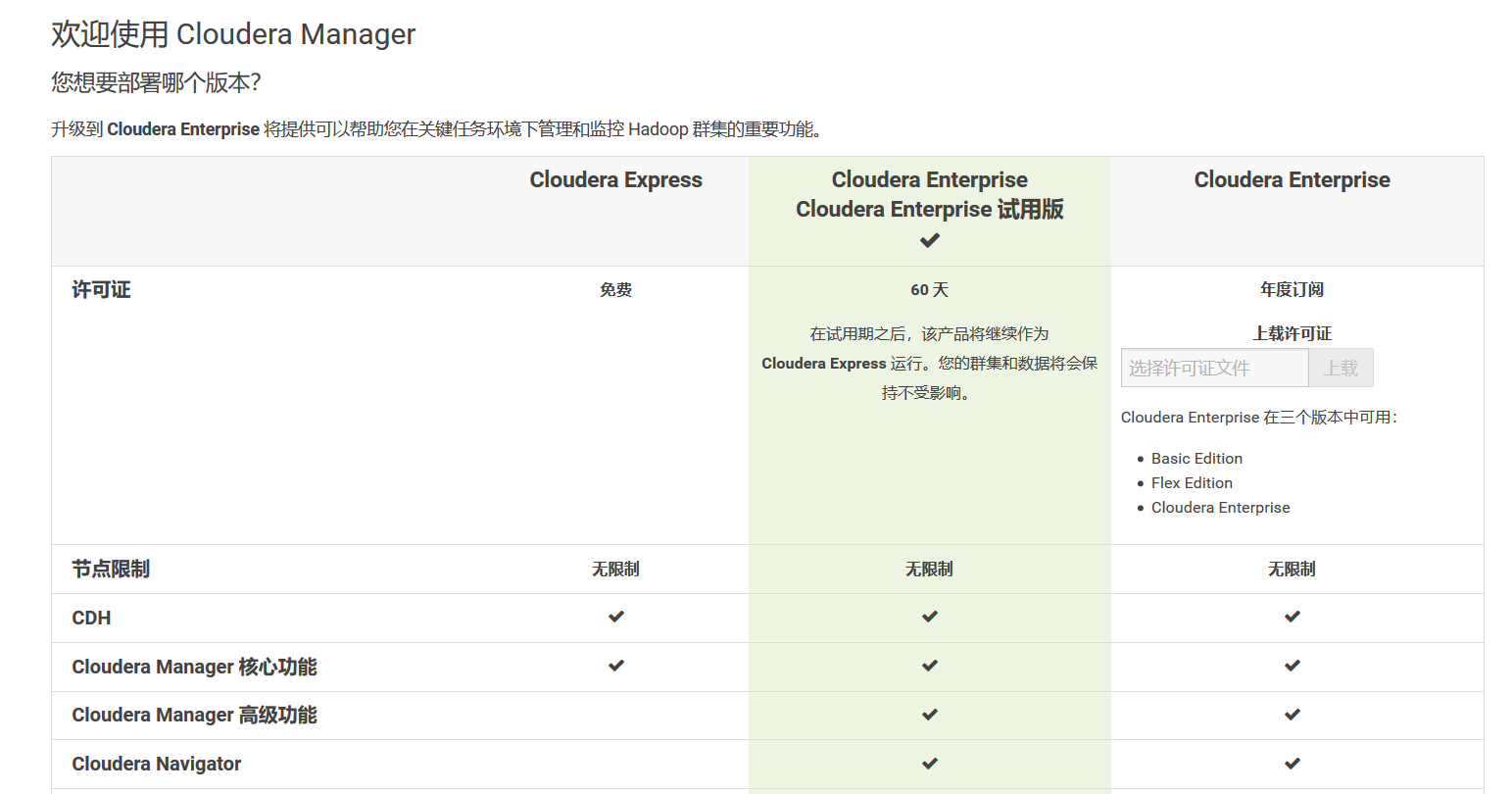
### 7、Cloudera Manager 安装

通过浏览器访问主节点的7180端口，默认的用户名和密码均为admin。

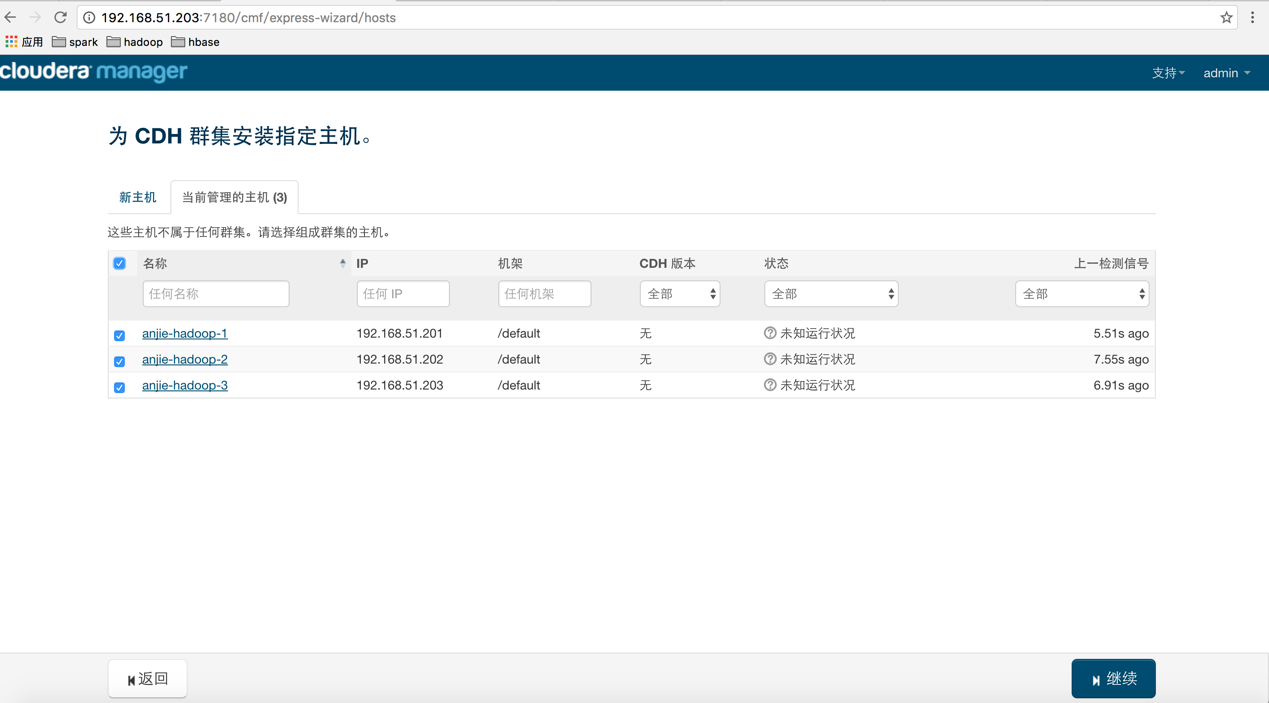
1)  登录地址：<http://192.168.2.100:7180>



2）选择cloudera free版本 或者试用版都可以



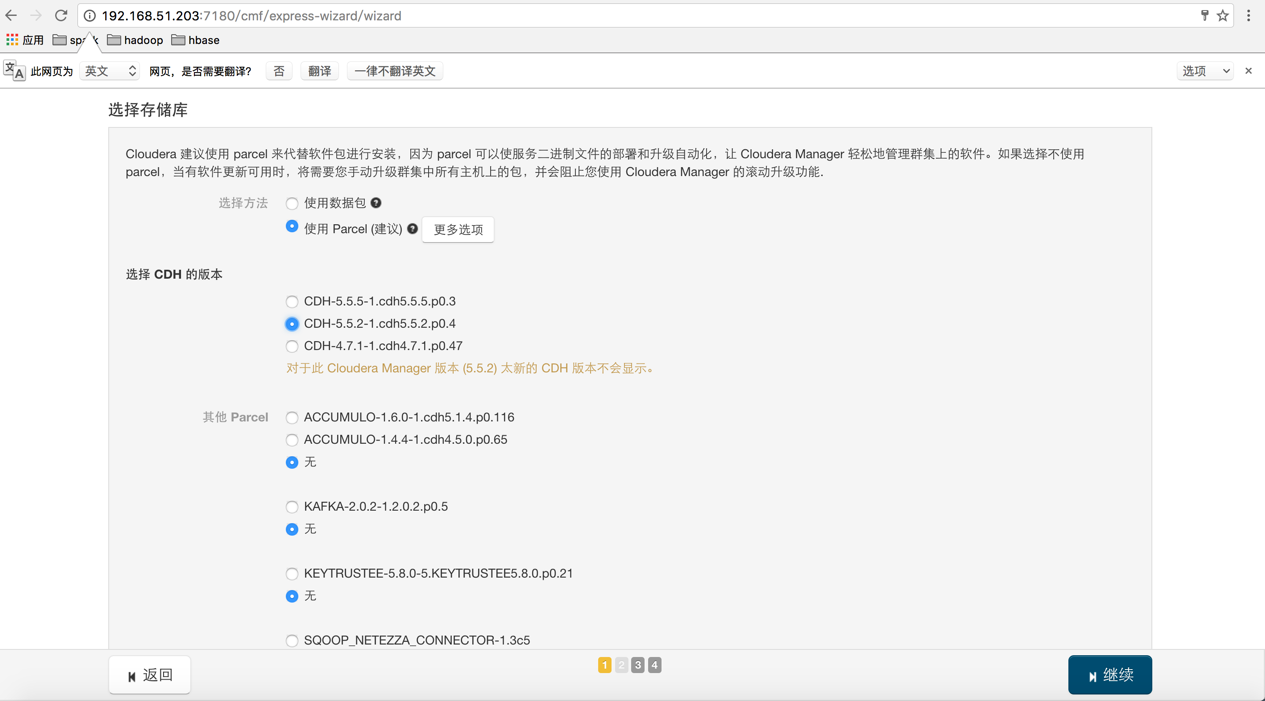
3）选择主机



也可以搜索主机进行添加



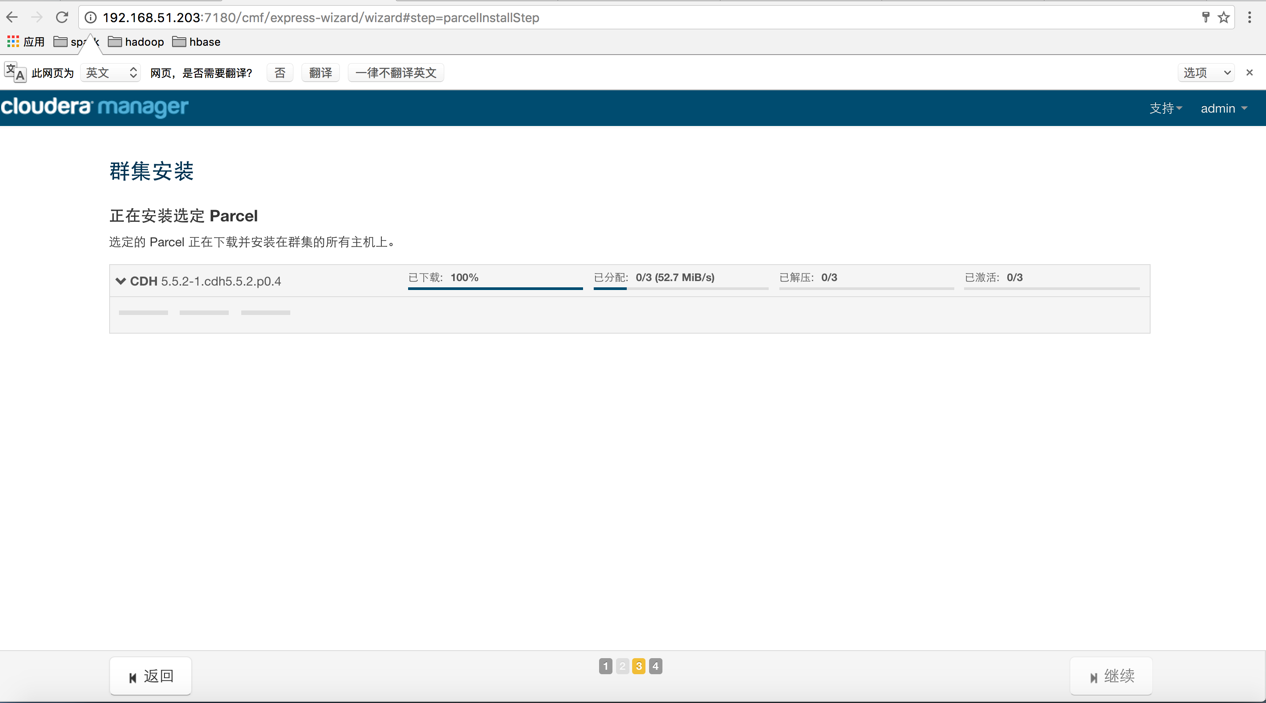
4） 选择存储库



如果修改了parcel\_repo的目录位置(/opt/cloudera/parcel\_repo)，需要在”更多选项”中修改，并重启cloudera-scm-server：

./ cloudera-scm-server restart

5） 集群的安装



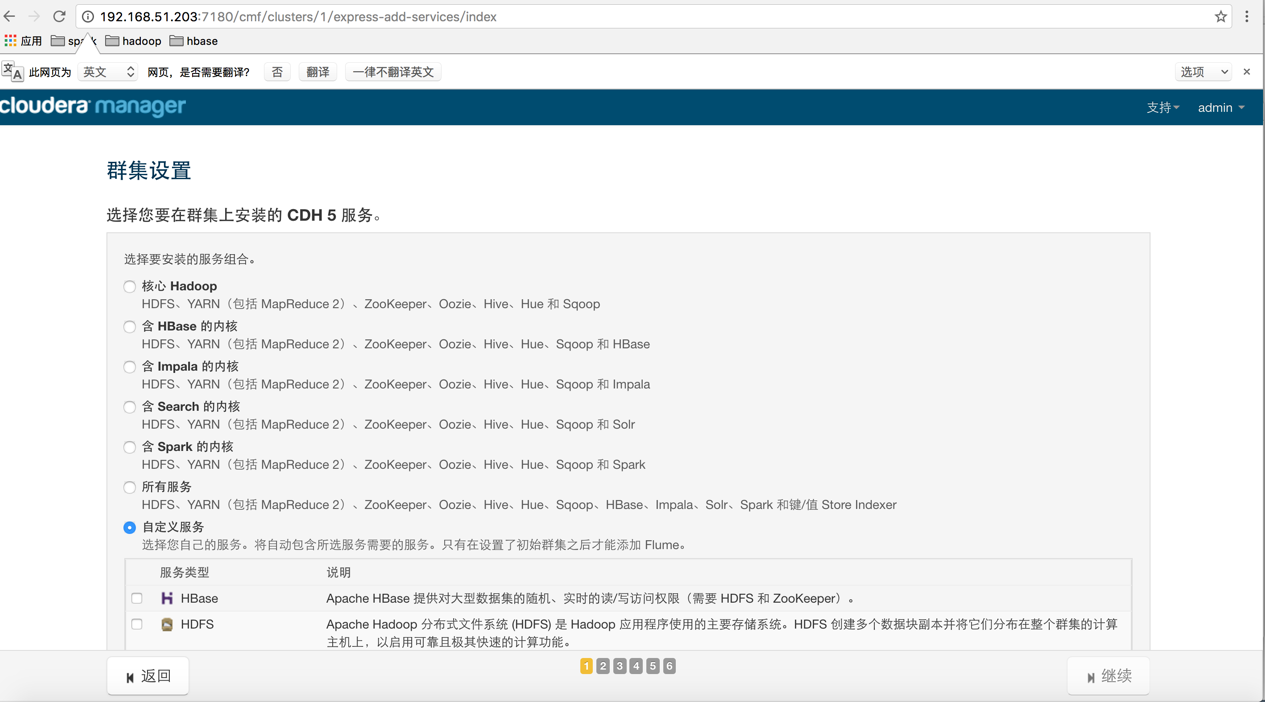
把master节点上的parcel\_repo下面的文件拷贝到其他的节点上，大概需要10-15分左右。



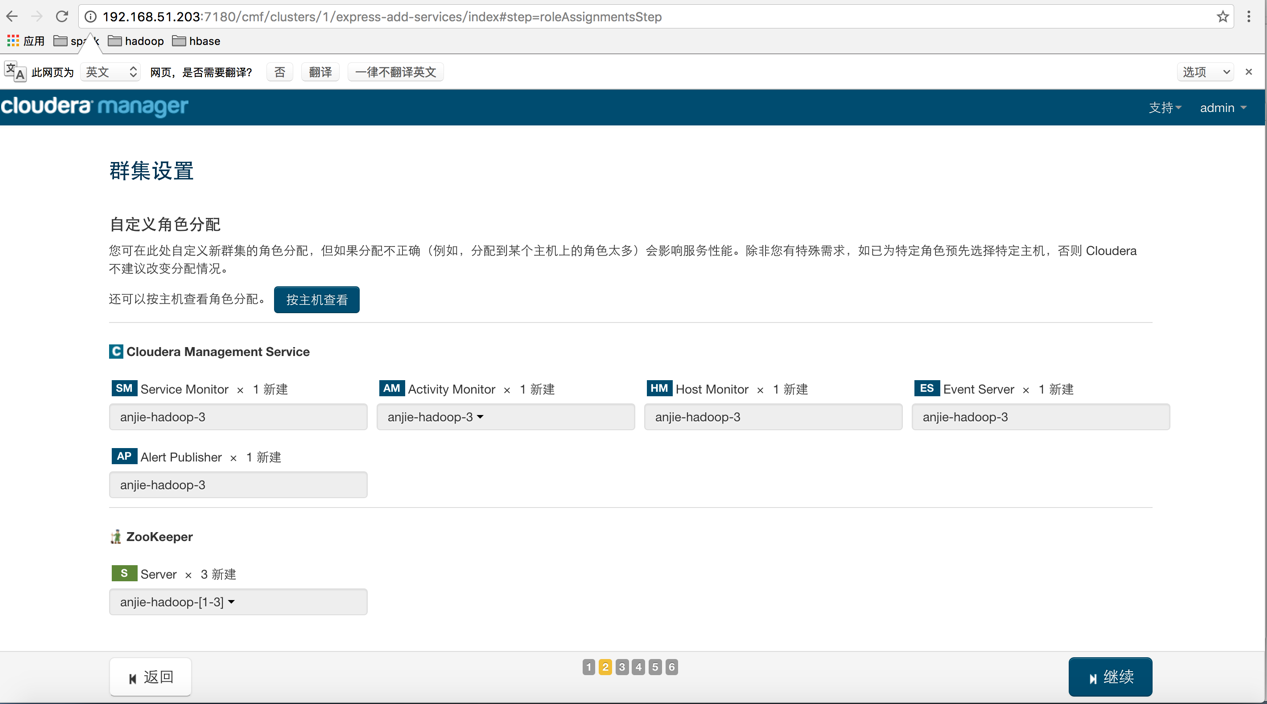
6）检查主机的信息



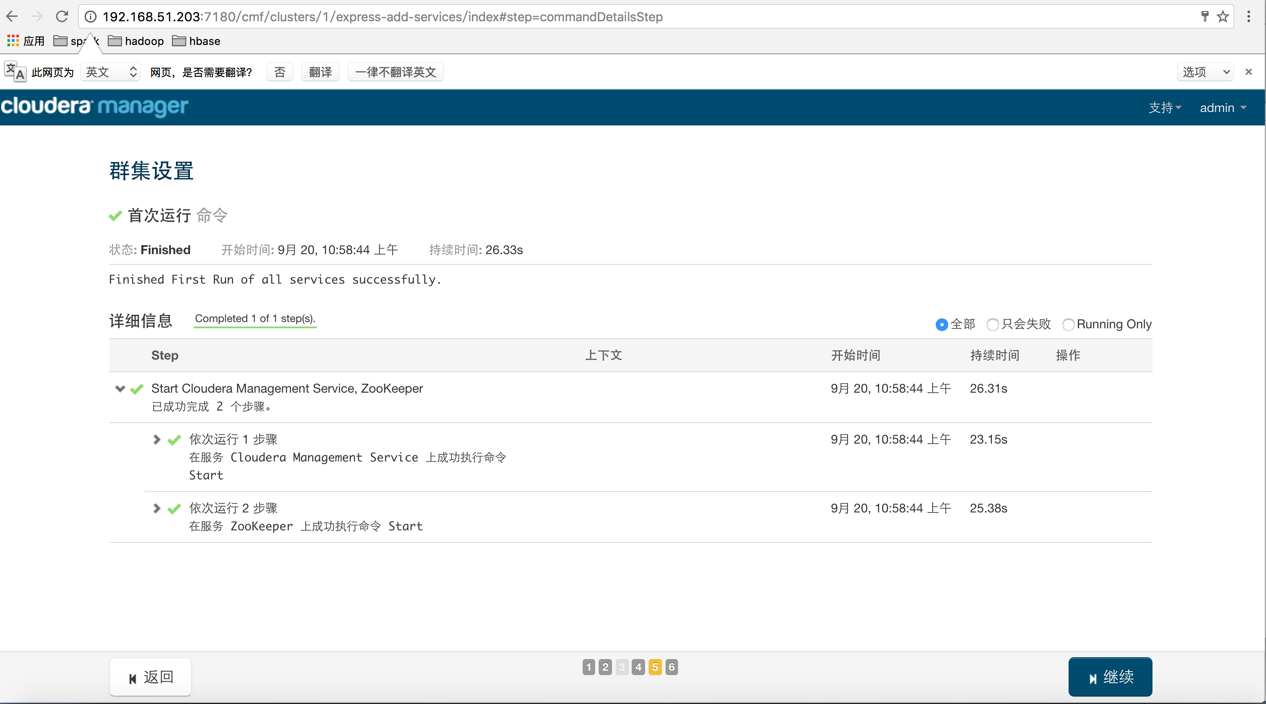
7）选择自定义安装



最小化安装，只安装zookeeper服务即可。



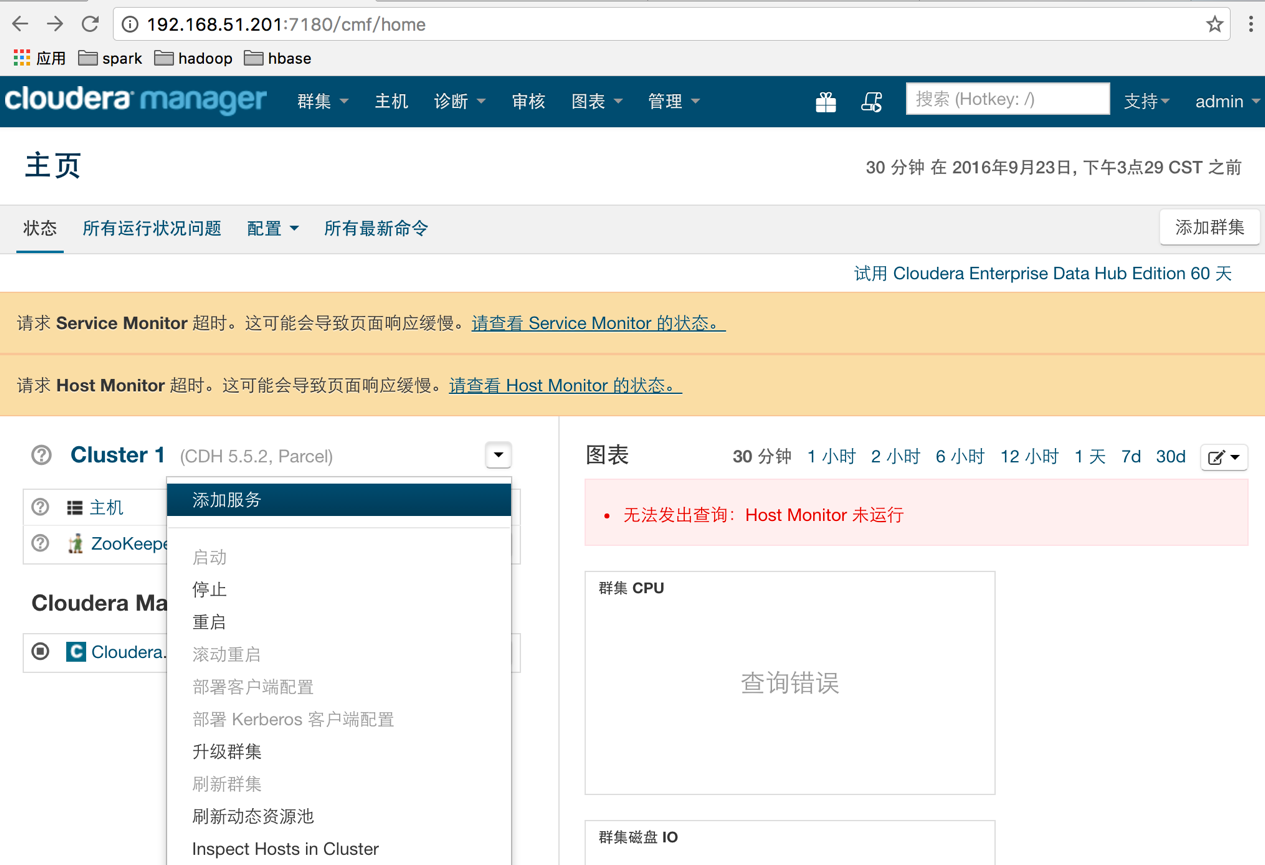
安装成功界面



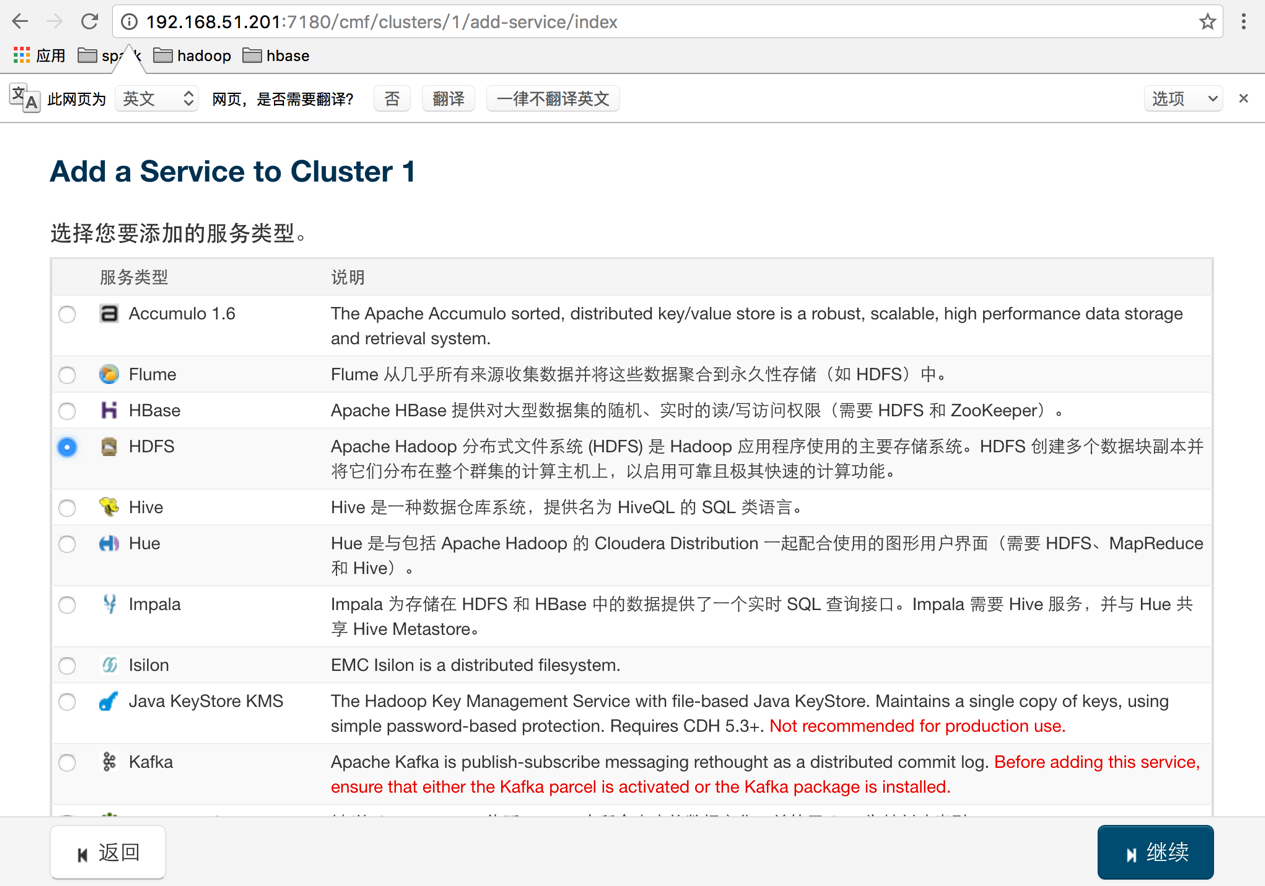
## 三、HDFS服务安装

### 1、HDFS安装

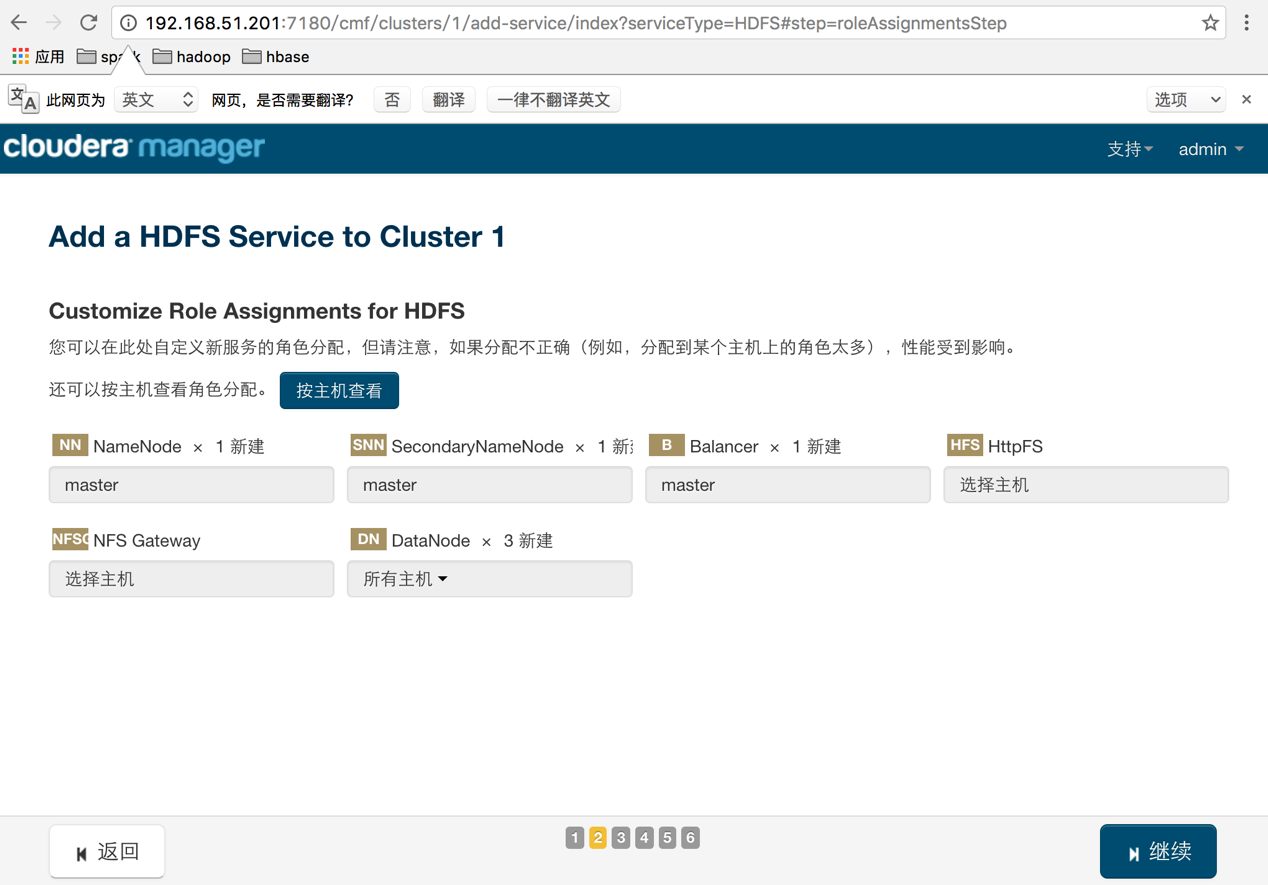
#### Step1: 点击添加集群服务按钮



#### Step2：选中要添加的服务，这里面我们选择HDFS服务。

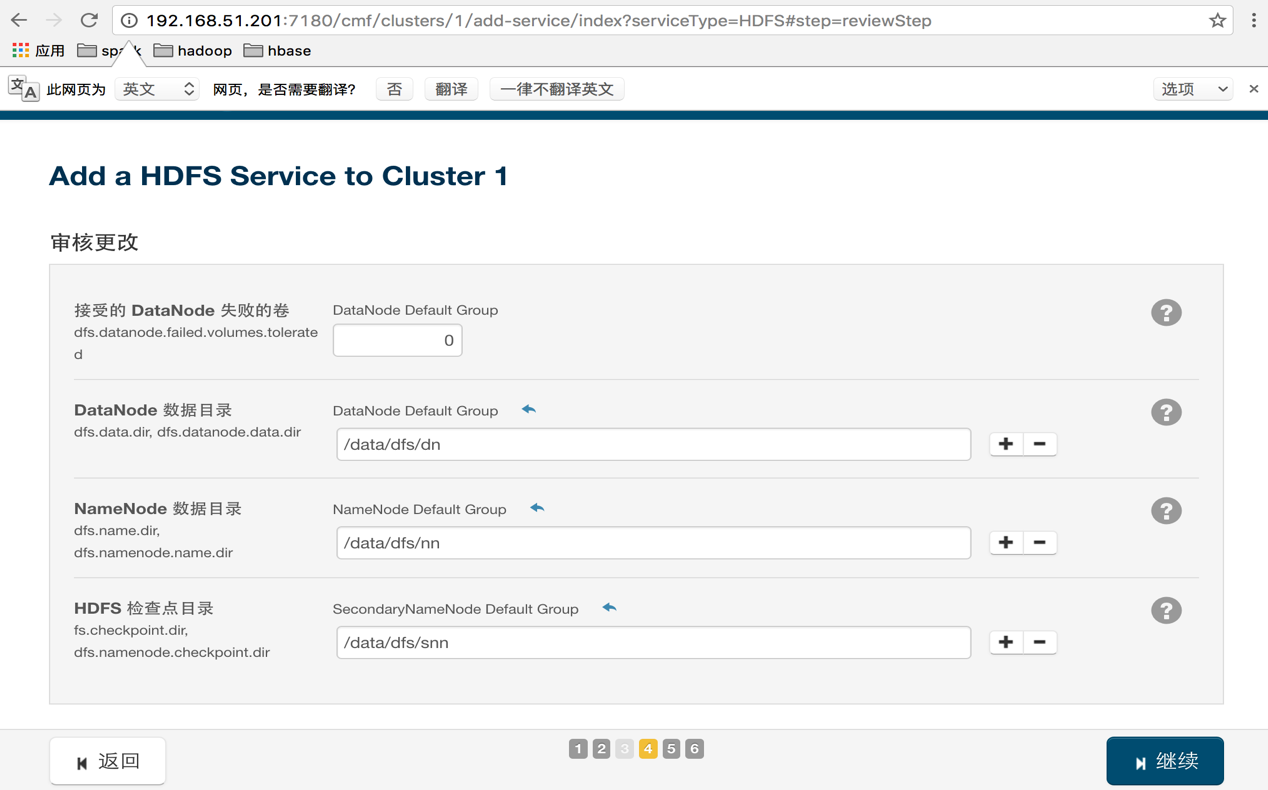


#### Step3：每台机器配置角色。



HDFS的服务的角色可以随便的配置，在此，我把namenode、secondaryNameNode和datanode安装在master节点，slave1、slave2节点安装datanode服务。

#### Step4：配置DataNode和Namenode数据存储的地方



次案例中，把datanode数据放在/data/dfs/dn下面，Namenode数据放在/data/dfs/nn下面，HDFS的checkpoint目录指定在/data/dfs/snn下面。

这些目录大家可以随便的修改，多个目录之间用逗号隔开。

#### Step5：HDFS服务的初始化和启动过程。



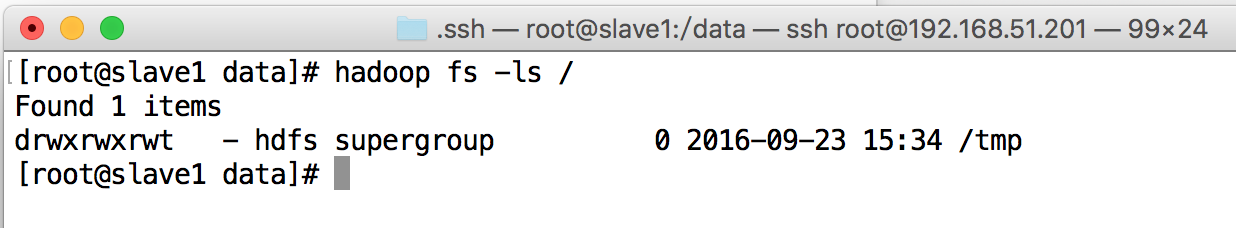
#### Step6：安装成功提示。



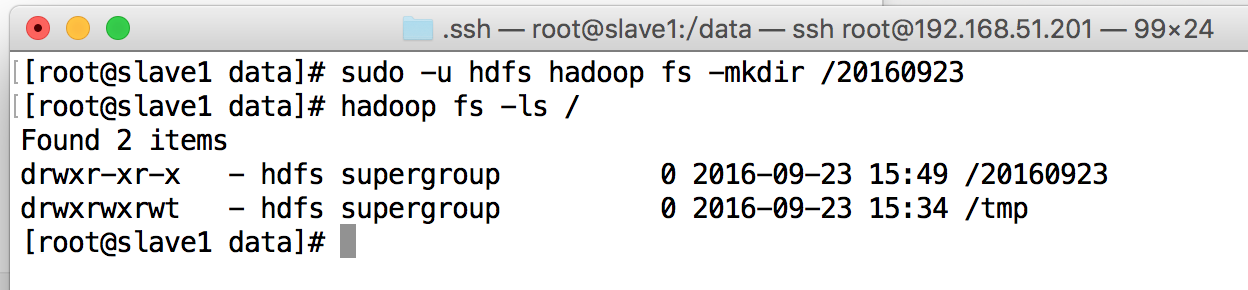
到此，HDFS的服务安装完成。

### 2、HDFS安装测试

#### Step1：查看HDFS的文件目录：



#### Step2:创建文件夹



恭喜您，HDFS成功安装成功！